7 Terahertz Spectroscopy of Biomolecules

Edwin J. Heilweil and David F. Plusquellic
National Institute of Standards and Technology

CONTENTS

7.1 INTRODUCTION

A better comprehension of biomolecular function and activity in vivo requires a detailed picture of biopolymer secondary and tertiary structures and their dynamical motions on a range of timescales.[1–3] At the core of all proteins, for example, are the substituent amino acid building blocks that determine their structure and function. In an effort to develop methodologies to directly monitor complex macromolecule dynamics in real time, we[4] and others[5,6] suggested that an atomic level picture of the concerted motions of polypeptide chains and DNAs may be accessible through accurate measurement of low-frequency vibrational spectra. These vibrations are expected to occur in the terahertz (THz) frequency regime and may be observed using Raman, low-energy neutron[7] and infrared absorption spectroscopies or related optical techniques. However, for even naturally occurring proteins and DNAs with >30 kDalton molecular weights and large numbers of constituent peptide units or bases, one would expect the density

269

of overlapping states to be so high in this frequency range that contributing absorption bands would "smear out" and yield essentially structureless spectra.

It would clearly be of highest priority to obtain low-frequency vibrationally resolved spectra for biologic systems in aqueous-phase environments because this condition would most closely mimic their natural environment. However, this scenario has not been immediately feasible for most systems using far-infrared absorption spectroscopy because (1) absorption by the amino acids and most other biomolecules is masked by much stronger water absorption in the 1–3 THz spectral region[8] and (2) spectral broadening arising from the full accessibility to conformational space and the rapid time scale for interconversion in these environments. Despite this limitation, recent studies of biomolecules in aqueous or high relative humidity environments reviewed in this chapter have revealed detailed information about the dynamics of these systems through careful broadband absorption measurements. Also reviewed here is the work performed to determine whether pure solid samples of organic model systems and protein fragments do indeed show sharp spectral features that are uniquely determined from their individual molecular symmetries and structure. Spectral THz absorption data and spectral modeling methods discussed in this review are the first of their kind and demonstrate that these expectations are indeed borne out.

7.2   EXPERIMENTAL PROCEDURES

As discussed in detail in our previous publications, solid organic species (e.g., di-substituted benzenes) and lyophilized samples of individual purified amino acids and short polypeptides were used as received from Sigma-Aldrich, Inc., or ICN Biomedicals, Inc.[9] Powders requiring low-temperature storage were maintained at 273 K until use to prevent decomposition and exposure to atmospheric water. Matrix diluted samples for THz absorption measurements were rapidly prepared by first weighing 2 to 10 mg of solid and homogenizing the material in a mortar and pestle to reduce the solid particle size distribution. This procedure ensures particle sizes sufficiently smaller than THz wavelengths to reduce baseline offsets at higher frequencies arising from nonresonant light scattering. Each sample was thoroughly mixed with approximately 100 mg of spectrophotometric grade high density polyethylene powder (Sigma-Aldrich, Inc. with <100 micron particle size, or MicroPowders, Inc. with <10 micron particle size) and pressed as a pellet in a 13-mm diameter vacuum die at the lowest possible pressures (approximately 200 psi or 1.4×10^6 Pa) to minimize decomposition from transient heating or high pressure. The 2.5-mm thick pellets have sufficient pathlength to eliminate etaloning artifacts (approximately 2 cm^{-1} period) often observed in THz spectra. Sample pellets were mounted in an aluminum sample holder fixed in position with a Teflon securing ring for 298 K measurements. Pellets were also encapsulated in a brass fixture that adapts to the copper cold finger of a vacuum cryostat (Janis Research Company, Inc., model ST-100)[9] fitted with 3-mm thick high-density polyethylene windows for broadband infrared studies at 4, 10 or 77 K.

Infrared absorption spectra in the 0.2–22 THz range were obtained with a modified Nicolet Magna 550 Fourier Transform infrared spectrometer (FTIR) using a

silicon-coated broadband beam-splitter and DTGS room temperature detector fitted with a high density polyethylene window. The dry-air purged, globar source FTIR has sufficient sensitivity to generate high quality spectra in the 1.2–22 THz (40–700 cm^{-1}) range for both room temperature pellets and samples placed in the liquid nitrogen or helium cooled cryostat positioned in the sample compartment. All measurements made using the FTIR (after ~1 hour of sample compartment purging to eliminate water vapor interference) were obtained at 4 cm^{-1} spectral resolution and averaging 64 interferometric scans. In standard fashion, spectra were converted to optical density (OD) units after ratioing raw sample transmission spectra (T_{sample}) to that of a pressed 100 mg polyethylene blank (T_{PE}) pellet (OD = $-\log_{10}(T_{sample}/T_{PE})$) obtained under identical acquisition conditions. Absorption band intensities were also found to obey Beer's Law (OD = εcl where ε is the molar extinction coefficient, c the molar concentration, and l the sample thickness) as the amount of investigated solid is varied in several pellet samples of the same thickness and weight. This finding strongly indicates that all observed absorption features directly arise from the sample in the host matrix rather than from pellet mixing inhomogeneity or other optical anomalies.

Lower frequency THz time-domain spectroscopy (THz-TDS) was used to collect spectra between 0.4 and 8 THz (10 cm^{-1} and 225 cm^{-1}) for solid disks and 10 to 90 cm^{-1} for liquid samples and compared with the FTIR spectra whenever possible. The THz-TDS spectrometer we employ is now a common apparatus and a review of the technique and set-up are reported elsewhere.[10] Details and modifications made to our instrument to extend the typical 0.3 to 3 THz spectral bandwidth out to approximately 8 THz are described as follows. Broadband THz pulses were generated from amplified 800 nm fs laser pulses (Ti:sapphire MIRA oscillator and Legend amplifier system, Coherent Lasers, Inc., with ~45 fs pulse duration, 1 kHz repetition rate, and an average power of 300 mW) by difference frequency mixing in a 0.150-mm thick GaP(110) crystal. Four 90° off-axis parabolic aluminum coated mirrors were used to propagate, focus, and collect the THz pulses to interrogate the sample in transmission mode. Transmitted THz pulses were then focused onto a second 0.150 mm GaP(110) detector crystal gated by weak (~200 nJ) 800-nm pulses for electrooptic detection. These weaker gate pulses were generated from the original input 800-nm pulses by reflection from the front surface of a quartz wedge before reaching the generation crystal. Linearly polarized gate pulses, modulated by the instantaneous THz field in the second 0.150-mm thick GaP detector crystal, were subsequently detected by balanced matched silicon detectors, and time-domain data was collected by a lock-in amplifier (Stanford Research System Model 830). The analog time domain data were transferred to a computer, digitized, and analyzed by a LabView interface and fast-Fourier transformed to obtain the THz power spectrum. In similar fashion to FTIR measurements described previously, raw sample spectra are corrected by using a PE pellet background spectrum to obtain sample absorption spectra. FTIR and THz-TDS spectra were collected after purging each instrument to obtain a dry air environment and thereby remove gas phase water interference. Analysis for comparison of averaged spectra (512 scans for FTIR and 20 for THz-TDS measurements) to theory yielded an intensity uncertainty from the baseline of less than ±0.005 optical density (OD) units (k = 1; type B analysis).

High-resolution THz absorption spectra were obtained using an instrument described in detail elsewhere.[11] Briefly, THz radiation from 0.06 to 4.3 THz was generated by the photomixing of two near IR laser at the surface of a low-temperature-grown GaAs photomixer.[12] The two lasers include a fixed frequency diode laser operating near 850 nm ($\Delta\nu_{FWHM} \approx 0.0001$ cm^{-1}) and a standing-wave Ti:sapphire laser having a resolution of $\Delta\nu_{FWHM} \approx 0.04$ cm^{-1}. The Ti:sapphire beam is combined with the output from a fixed frequency diode laser and focused by an aspherical lens onto the photomixer. A maximum power of ≈ 1 μW of THz radiation is obtained from 0.06 THz to 3 THz. The focused THz beam passes through the cryogenically cooled sample at 4.2 K and is detected by a liquid helium–cooled silicon-composite bolometer. Power detection sensitivity of the bolometer is <1 nW up to 3 THz in a 400 Hz bandpass (NEP of the bolometer is 1 pW/Hz$^{1/2}$).

7.3 THEORETICAL SPECTRAL MODELING METHODS

To better understand the origin and assign the absorption features observed in our experimental spectra to molecular motions, a complementary series of calculations were performed to try to extract a picture of the intramolecular and intermolecular vibrational modes for these samples. Obtaining both accurate vibrational mode frequencies from properly minimized structures and extracting infrared intensities throughout the THz spectral region is crucial when comparing experimental observations to theory.

Gaussian 03 Calculations: To model THz spectra of minimally interacting (no hydrogen bonding) model systems such as dicyanobenzene, gas phase density functional geometry optimizations, and vibrational frequency calculations for each molecule was performed by the Gaussian software package (G03W Rev B.02)[13] with the Becke-3-Lee-Yang-Parr (B3LYP) functional and 6–311++G* basis set under the Windows 32 operating system environment. Initial starting atomic geometries for each molecule were obtained from the NIST chemical structure database,[14] and the structures optimized as free space molecules. Positive vibrational frequencies and close agreement of bond lengths and angles with published crystallographic data for the systems examined confirmed that the minimized structures were obtained.

CHARMm Force Field Calculations for amino acids: Calculations of the THz vibrational spectrum of l-serine, l-cysteine, l-tryptophan, and dl-tryptophan were accomplished using the CHARMm 22 empirical force field.[11,15] The method used is similar to the previous work of Gregurick et al. for the calculation of the vibrational spectrum of small peptide[16] and sugar molecules.[17] For the amino acids, crystallographically determined atomic coordinates were used as the starting structure for the calculations.[18–23] To represent the crystal environment of the experimental work, we constructed a pseudo–crystal lattice comprising single-unit cells; the macroscopic crystal was then reproduced by using periodic boundary conditions in all calculations. We first optimized the geometry of the crystals using an Adopted Newton Raphson method until the gradient change was less than 10^{-4} kcal/mol. Using the steepest descent method, the structure was gently optimized further to

ensure an energy minimum, whereby all eigenvalues of the mass-weighted Hessian were positive.

The THz vibrational frequencies were obtained from the normal mode eigen-values of the mass-weighted Hessian and the intensities of each normal mode were calculated from their dipole derivatives. The dipole derivatives were cal-culated numerically from the displacement of the equilibrium structure along each normal mode direction. Computed spectra were represented by taking the calculated intensity for each absorption peak convoluted with a Gaussian spectral function with a FWHM corresponding to 20 cm[–1]. All calculated spec-tra were strictly obtained at the harmonic level and the calculations were per-formed on a UNIX-based Silicon-Graphics dual processor system taking about 15–30 minutes per simulation.

DFT Spectral Analyses: DFT calculations on the crystal structures of selected amino acids using the freeware program CPMD, version 3.9.2.[24] For l-serine, the cal-culations used the crystal structure coordinates.[18–23] However, to save CPU time, the l-cysteine amino acid calculations used the minimized structure output of CHARMm described previously. For these calculations, a super cell was created with four mol-ecules per cell and then applied periodic boundary conditions to simulate the larger crystal lattice. The one-electron orbitals were expanded in a plane wave basis set with a kinetic energy cutoff of 80 Ry restricted to the gamma point of the Brillouin zone. Medium soft norm-conserving pseudo potentials of Martins-Trouiller were used for all the elements.[25] The energy expectation values were computed in recipro-cal space using the Kleinman-Bylander transformation.[26]

To calculate the intensities of each normal mode, we allowed the system to propagate for a total time of 5 ps by applying ab initio molecular dynamics using 600 a.u. for the fictitious mass of the electron. In these cases, a time step of 5 a.u. (0.12 fs) was used and recorded positions and charges of each atom at every step. The total dipole moment was calculated at each step of the ab initio molecular dynamics using the Berry phase polarizability.[27] IR intensities were computed from the auto correlation of the dipole magnitude followed by taking the Fourier trans-form.[28] Again, each spectral peak was represented as a Gaussian bandshape with a FWHM corresponding to 20 cm[–1], which closely mimics the observed experimen-tal widths. CPMD calculations were performed on the NIST multinode LINUX cluster using 4–8 nodes and took approximately 2–3 weeks for the optimization and frequency calculations. In contrast, the molecular dynamics simulations used to extract dipole derivative intensities required several months of computation for each species.

DFT calculation on crystalline tripeptides were performed using the DMol[3] soft-ware package.[29] All calculations included all electrons and made use of a double numerical plus polarization atomic orbital basis set (equivalent to a double-ζ). The exchange-correlation functional based on the generalized gradient approximation in the parametrization of Perdew and Wang (PW91)[30] was selected for its known reli-ability to accurately model hydrogen bonded systems. The crystal cell parameters of the monoclinic cells were held fixed at the values reported from the X-ray studies. The harmonic frequencies and normal mode displacements were calculated at the

gamma point by a two-point numerical differentiation of the forces. The intensity for each normal mode, Q_i, was calculated using

$$I_i \propto \left| \left\langle \upsilon' \left| \frac{\partial \mu_i}{\partial Q_i} \right| \upsilon'' \right\rangle \right|^2 \cong \left| \sum_j^{3N} \frac{e_j q_{i,j} - e_j^o q_{i,j}^o}{\Delta Q_i} \right|^2 \cong \left| \sum_j^{3N} \frac{e_j^o \left(q_{i,j} - q_{i,j}^o \right)}{\Delta Q_i} \right|^2$$

where the sum is overall 3N Cartesian coordinate displacements, $q_j - q_j^o$, derived from the eigenvectors of the Hessian by dividing by square-root of the atomic masses and ΔQ_i represents the root-mean-square displacement of atoms in mode Q_i. The static charges, e_j^o, were set to the Mulliken atomic charges of the optimized structure.

7.4   WEAKLY INTERACTING ORGANIC MODEL COMPOUNDS

In weakly associated solid-state systems, most intermolecular interactions are dominated by noncovalent, electrostatic, or Van der Waals forces. To explore whether THz spectra of weakly interacting planar molecular solids (crystals) and analogous solution phase systems can be modeled using available force fields and the detailed molecular interactions in this spectral region, we measured and calculated the THz spectra of the 1,2-, 1,3-, and 1,4- isomers of dicyanobenzene (DCB) and related dinitrobenzene structures which differ only by the ring location of the two cyano (or nitro) side groups.

Vibrational THz spectroscopy and mode assignment investigations of the DCB isomers was previously conducted,[31] but little attention was paid to the lower frequency THz region of the spectrum. Because the DCB homologs are known to have planar, weakly interacting structures in the solid state,[31] investigation of the DCB isomers could reveal whether available cost-effective isolated gas phase molecule spectral calculations using density functional theory (DFT) adequately agree with experimental THz spectra. In addition, having strongly coupled electronegative cyano groups coordinated to the pi-bonded benzene ring affords the possibility that partial atomic charges undergoing vibrational dipolar motions could enhance THz absorption intensities making comparison to theory straightforward. Many of these experimental and theoretical characteristics were confirmed and briefly reviewed here.

To first give the reader a flavor for how many THz modes involve motion of the constituent atoms, the vector representations of the calculated vibrational modes for frequencies below 200 cm^{-1} of each molecule are shown in Figure 7.1. 1,2-DCB and 1,3-DCB exhibit three IR-active modes in this spectral range but 1,4-DCB has only two IR-active modes due to inversion symmetry.

The calculated number of frequencies and intensity profiles match very well with the full experimental spectral observations. Figure 7.2 presents the collection of experimental and modeled vibrational spectra from the gas-phase Gaussian calculations, FTIR and THz-TDS for solutions (in CHCl$_3$) and solids (in PE matrix) measurements for all three isomers at room temperature (ca. 298 K). As one can see, the calculated spectral frequencies and relative intensities agree extremely well for all three isomers. There appears to be no well-defined frequency shift pattern one might apply for the calculated spectra to exactly overlap the observed features.
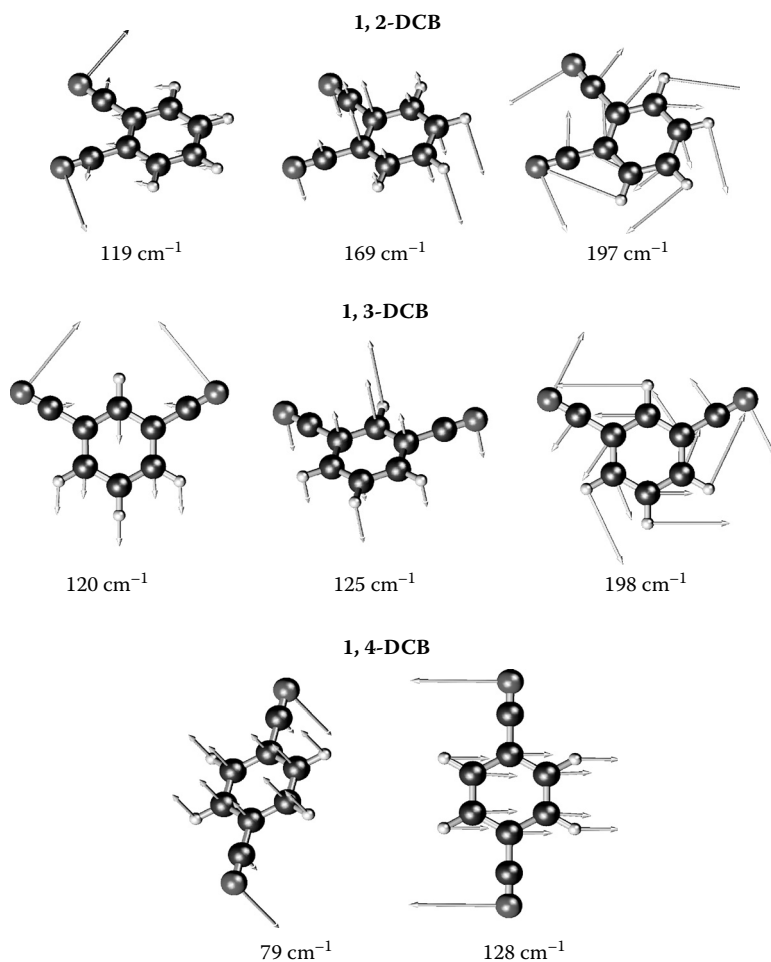
1, 2-DCB

119 cm^{-1} 169 cm^{-1} 197 cm^{-1}

1, 3-DCB

120 cm^{-1} 125 cm^{-1} 198 cm^{-1}

1, 4-DCB

79 cm^{-1} 128 cm^{-1}

FIGURE 7.1 Displacement vector representation of the lowest non-zero infrared-active modes of the 1,2- 1,3-, and 1,4-dicyanobenzene isomers obtained from Gaussian 03 vibrational analysis of optimized isolated gas phase structures.

Comparing the THz-TDS and FTIR measurements of solid and liquid-phase samples (eliminating intermolecular modes) shows that these systems exhibit two intense IR-active intermolecular (phonon) features for all three species in the spectral region below 100 cm^{-1}.

This study yielded two significant observations: (1) The Gaussian software package may be reliably used to interpret low frequency THz spectral region for weakly interacting organic systems and it is able to help distinguish and classify observed vibrational features as intramolecular or intermolecular (phonon) modes. This study and our experience with similar molecular structures (e.g., dinitrobenzenes) also shows that DFT with B3LYP level and 6–311G* or higher order basis adequately
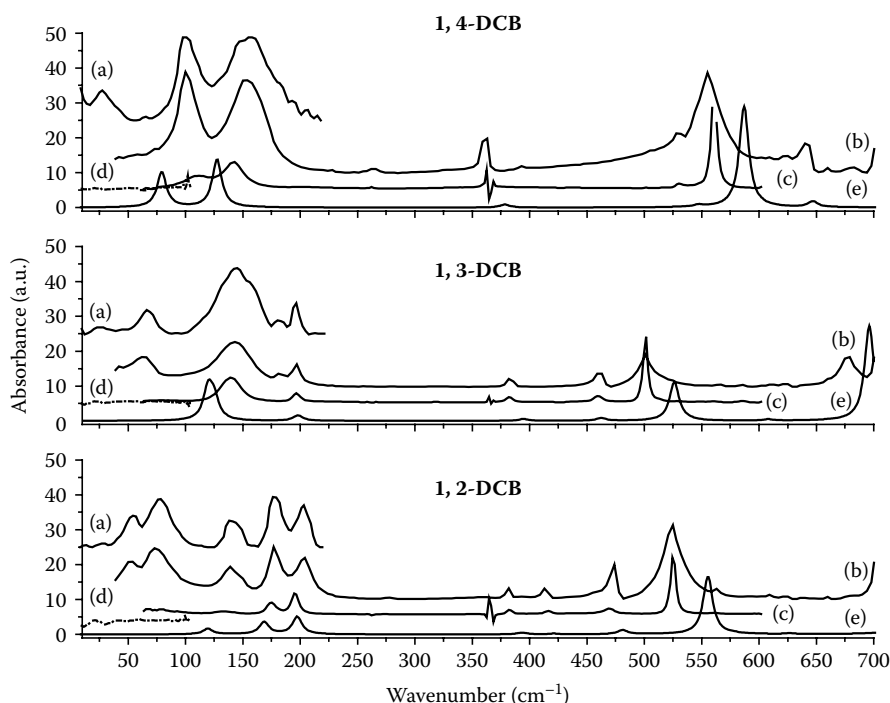
FIGURE 7.2 Experimental and simulation spectra of DCB isomers. Spectra of solid phase in PE matrix measured by (a) THz-TDS and (b) FTIR, of solution phase (in chloroform) measured by (c) FTIR and (d) THz-TDS, and (e) DFT simulations. To simplify, solution spectral intensities are multiplied by 10, while the solid THz spectra are scaled by 2. Spectra are vertically offset for clarity. See Discussion for extinction coefficients for the isomers.

determines the solid-state equilibrium structure and internal modes with reasonably good accuracy for systems exhibiting weak intermolecular interactions. (2) To the best of our knowledge, the observed intensities of the lowest two phonon modes of 1,4-DCB are the strongest THz absorption features reported to date. These isomers and their readily identifiable spectral features may be of great interest in future studies using THz spectroscopy to analyze internal and phonon mode properties of solids.[32]

7.5 SMALL BIOMOLECULES AS CRYSTALLINE SOLIDS

The THz spectra of biomolecular crystalline solids have been reported for a variety of oligopeptides, simple sugars, pharmaceutical species, and other small biomolecules. In contrast to the larger systems discussed in later sections, a common feature of these studies is the observation of relatively small number of sharp vibrational features attributed to a discrete number of intramolecular or phonon modes of the solid. Observing vibrationally resolved features is primarily a result of two factors. Depending on the space group, typically only one or two conformations of the

monomer subunits are dominant in the crystal and therefore, the small number of unique atoms per unit cell results in a small number of vibrational modes that are IR active below 3 THz. Second, when employing the proper crystal growth procedures that minimize defects, narrow inhomogeneous line widths (<1 cm^{-1}) often result especially for systems where strong hydrogen bonding networks exist throughout the solid. Moreover, at lower temperatures approaching 4 K, the spectral lines can significantly sharpen and shift depending on the impact of anharmonicity and crystal cage effects. Temperature dependent studies can give information about the character of the mode as well as the anharmonicity of the potential energy surface. A few examples in each of these areas will be reviewed.

7.5.1   AMINO ACIDS

In an effort to reveal the underlying THz spectral contributions from the individual amino acids of a protein, we began a series of experiments to measure the THz IR absorption spectra for many of the 20 naturally occurring solid-state (lyophilized) amino acids. As discussed, we found that these spectra demonstrate a high degree of complex absorption features and structural information in the 1–22 THz (33 cm^{-1} to 700 cm^{-1}) spectral range.[33] These results complement a related investigation of short-chain polypeptides also published by our group and discussed later in this chapter.

We obtained the solid-state spectra for l-serine and l-cysteine to determine whether an atomic substitution (–OH for serine versus –SH for cysteine) can be identified in the THz spectrum.[34] Significant differences were noted, especially in the lower frequency phonon region (<100 cm^{-1}) indicating the crystalline lattice interactions are strongly affected by the simple atomic substitution.

THz spectroscopy is also highly sensitive to the enantiomeric crystalline structure of biomolecules. The stereochemistry of many biologic and pharmaceutical species plays a significant role in whether it is functional or even toxic. To this end, an investigation into whether the THz spectrum of enantiomeric mixtures of left- and right-handed species can be uniquely identified was undertaken by our group.[35] Spectra for l-, d- and dl-tryptophan (1:1 l- and d-tryptophan) as well as various mole percent mixtures were studied using broadband FTIR spectroscopy. Although the spectra for pure d- and l-tryptophan are identical, admixtures from about 5 to 50 percent mole fraction could readily be identified from changes in absorption intensities of several spectral features. Figure 7.3 shows examples of these spectra as well as a single fluorine-substituted analog which generates a completely different spectrum compared with the native amino acid. Principle component spectral analysis could readily be employed to quickly analyze these spectra and readily extract "contaminant" concentrations, a task that is difficult for any other analytical method.

Starting from available X-ray crystal structures for each amino acids species,[14] application of inexpensive DFT methods[13] did not produce modeled THz spectra that agreed well with experiment. Apparently the currently available basis sets or parameterized potential energy functions are not adequate to model intermolecular hydrogen bonding in these systems. It was suggested that application of modern ab initio structure, frequency, and intensity calculations will eventually enable assignment of THz spectral features to both internal and external vibrational motions of
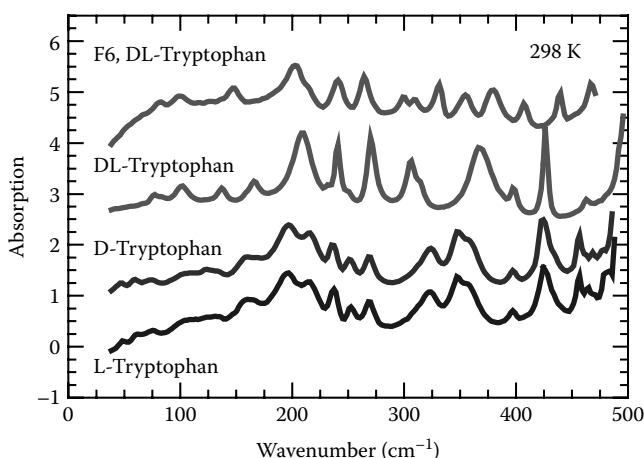
FIGURE 7.3 Experimental infrared absorption spectra for the pure d- and l-tryptophan (Trp) enantiomers, their 1:1 mixture (dl-Tryptophan) and F6-fluorine substituted dl-tryptophan amino acid in the solid state. Comparison of the l-Trp and d-Trp stereoisomers shows the spectra are identical, whereas the dl-Trp mixture has a distinctly different spectrum. The F6-dl-Trp species also shows different spectral features than the dl-Trp mixture.

these complex hydrogen-bonded molecular solids. Application of CHARMm and CPMD approaches were found to agree better than simple gas-phase theory and these results were included in our publication.[34]

7.5.2 POLYPEPTIDES

Because polypeptides have C- and N-terminations and are therefore sequence specific, we investigated a series of polypeptides to examine whether their crystalline forms yield qualitatively different THz spectra. As may be seen in Figure 7.4, every peptide that was examined, including reverse sequences, produced a different and readily identifiable spectrum. From these results, it can be concluded that short-chain crystalline polypeptides may be uniquely "fingerprinted" by simply acquiring the THz spectrum in the 50 to 500 cm[−1] spectral range. Several mixed polypeptide sequences with up to three amino acids were also studied (e.g., Val-Gly-Gly, gluta-thione), but sharp spectral features were not present for species containing more than 10 amino acids. This observation may be the result of overlapping spectral density as the molecular weight increases, or the spectral features are swamped by hydrogen-bonded water, which is difficult to remove from these complex biomaterials.

The sensitivity of the THz region to different β-sheet structures and the impact of cocrystallized water was investigated in our group[36] for three different crystalline forms of the tripeptide, NH_3^+-Ala_3-O^-, using a continuous wave THz source and THz-TDS methods.[37] Depending on the conditions during recrystallization, trialanine is known from x-ray crystal work[38] to exist in one of two β-sheet forms, a dehydrated parallel β-sheet (p-Ala_3) and a hydrated anti-parallel β-sheet (ap-Ala_3-H_2O). A third form discovered in this study was the dehydrated form, ap-Ala_3. As shown in Figure 7.5, the
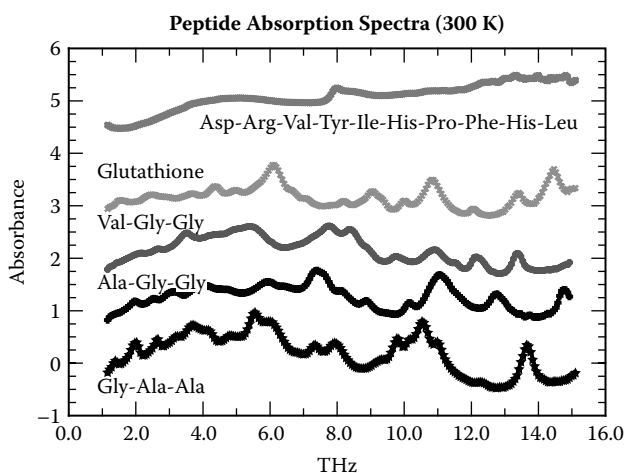
Peptide Absorption Spectra (300 K)

FIGURE 7.4 Representative THz infrared absorption spectra for several di- and tripeptides showing the variability in spectral content and uniqueness for each structure. For comparison (top), a random sequence 10-mer is also shown and exhibits loss of distinguishing absorptions, most likely from spectral overlap or hydrated water.
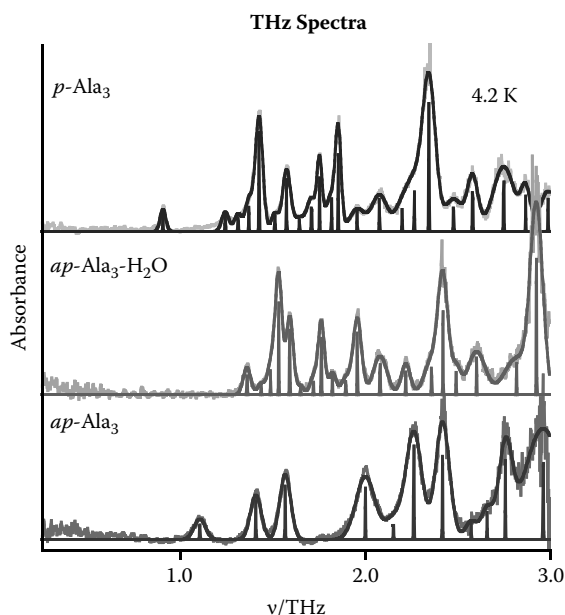
THz Spectra

FIGURE 7.5 THz absorption spectra of the three different crystalline forms of trialanine. Although the conformational forms of the monomers are similar, the parallel and hydrated antiparallel β-sheets have different space groups (P 12 and C 2, respectively, the dehydrated form is unknown but likely C 2). These spectra indicate the sensitivity to the different space groups as well as the impact of the hydrogen bonded (structural) water.

THz spectra (0.6 cm^{-1} to 100 cm^{-1}) obtained at 4.2 K for the three forms are qualitatively different. This is in sharp contrast to the mid-IR region where the FTIR spectrum[39] of ap-Ala$_3$-H$_2$O is similar to that of the p-Ala$_3$ and nearly identical to the dehydrated form, ap-Ala$_3$. The significance of this result is the dramatic impact the environment and the arrangement of hydrogen bonds has on the THz spectrum because the conformational forms of the monomers is essentially the same.

The small size of these systems enabled rigorous comparisons with quantum chemical (density functional theory) and classical models (CHARMm). In sharp contrast to the weakly bound organic systems discussed previously in which gas phase predictions were sufficient, calculations on these hydrogen-bonded crystals required the use of periodic boundary conditions. Results from the DFT calculations using the DMol^3 program suite[29] are shown together with experiment in Figure 7.6 for the parallel and hydrated antiparallel sheet forms. Spectral predictions of the parallel sheet with its highly cross-linked hydrogen bonding network are in fair agreement with many of the observed features. However, the weak intersheet binding in the antiparallel sheets is a severe challenge for theory where clearly the current level fails to achieve even qualitative agreement. Furthermore, the predictions from the classical model, CHARMm, in Figure 7.6 indicate that the similarities regarding spectral predictions were accidental because the corresponding nuclear motions at these two levels were vastly different.[36] These results suggest the need for improved density functionals that target multibody effects present in hydrogen bonding networks and improved force field models that include polarizability terms for water and three-atom hydrogen bonding terms for periodic solids. Current theoretical studies are under way that (1) use a triple zeta basis set and a fully relaxed crystalline cell geometry at the DFT level and (2) treat anharmonicity of the modes using VSCF theory at the classical level.[40]

AU: Please spell out VSCF

7.5.3 Nucleic Acid Bases and Other Sugars

The Jepsen group[41] studied the dielectric function of the four nucleobases, adenine (A), guanine (G), cytosine (C), and thymine (T) and their corresponding nucleosides, dA, dG, dC, and dT (d = dioxyribose) of DNA from 0.5 to 4.0 THz at 10 K and 300 K using THz-TDS. The spectra of A, G, C, and T are dominated by absorption features between 1 THz and 3.5 THz. At 10 K, the broad room temperature features blue shift by 5 percent and resolve into several narrow peaks. The blue shift with decreasing temperature is attributed to decreasing bond lengths. The spectra above 1.5 THz of dA, dG, dC, and dT are similar to the nucleobases but a second group of resonances from 1 THz to 2 THz have narrow asymmetric line shapes attributed to the sugar groups.

Quantum chemical calculation were performed on the gas phase tetramer of thymine using density functional theory (B3LYP/6-31G basis). The absorption spectrum and index of refraction were calculated for the minimum energy geometry (beginning with the X-ray crystal structure) and compared with experiment. The calculations predict four low frequency infrared-active modes arising from in-plane and out-of-plane intermolecular motions of the hydrogen bonded cluster. Although definitive assignments are not possible using this approach, nevertheless, the correct

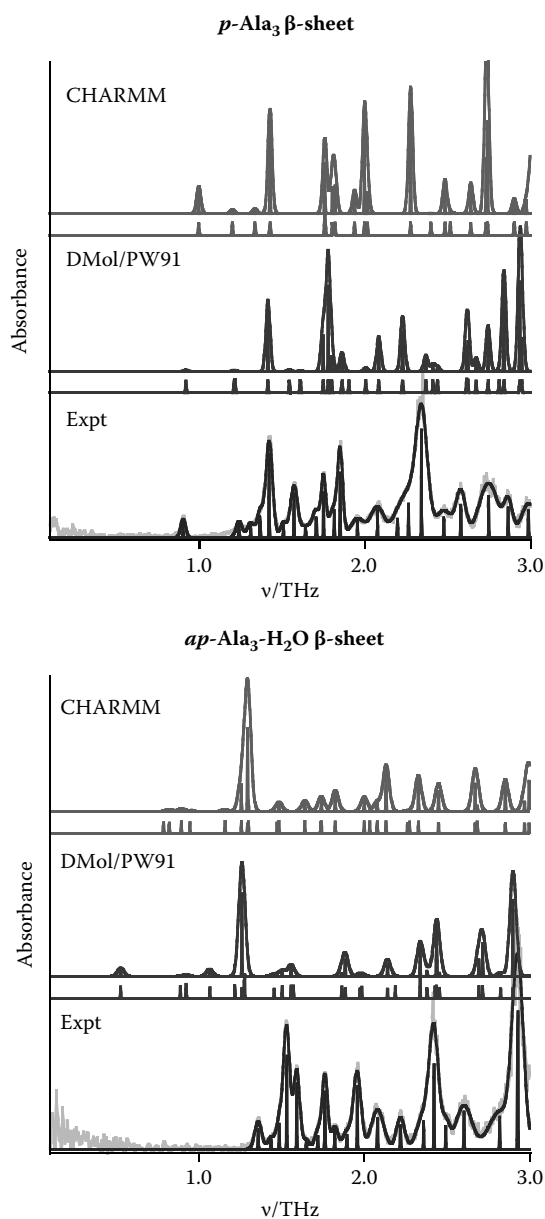**p-Ala$_3$ β-sheet**

ap-Ala$_3$-H$_2$O β-sheet

FIGURE 7.6 Predicted and observed THz spectra of the parallel (top) and hydrated antiparallel (bottom) β-sheet forms of crystalline trialanine from DFT/PW91 (DMol[3] software suite) and CHARMm.

number of features was reproduced. The significance of these results is the observation of the well-resolved and distinct spectral fingerprints of the nucleobases and nucleosides arising from the arrangement of the molecules in the microcrystalline structures.

The THz dielectric properties of the three sugars—glucose, fructose, and (protonated and deuterated) sucrose—were studied by the Jepsen group[42] from 0.5 THz to 4.0 THz using THz-TDS methods. Crystalline forms were shown to have sharp features (3 cm^{-1}), whereas absorption by the amorphous forms were featureless and increased monotonically with frequency. With increasing temperature, all absorption features of glucose and fructose broadened and redshifted. However, for sucrose, the two lowest frequency resonances of the protonated and deuterated samples (peak 1 at 1.38/1.28 THz for the protonated/deuterated forms and peak 2 at 1.69/1.63 THz) initially blueshifted with increasing temperature and then above a crossover temperature (240 K for peak 1 and 120 K for peak 2), shifted to lower energy. They found the frequency shifts with deuteration $\Delta \nu /\nu = -7$ percent and -3.5 percent are larger than expected based on simple harmonic mass weighting alone ($\Delta \nu /\nu \sim 1.2$ percent). Furthermore, they argue the anomalous blue shift is not easily explained based on vibrational anharmonicity because redshifts arising from hydrogen bonding stretch potentials are expected with increasing temperature. They suggest the initial blue shift at low temperatures may be a result of the relatively larger impact of van der Waals forces on the hydrogen bonding potentials compared with the higher temperature limit where anharmonicity presumably dominates.

7.5.4 OTHER SMALL BIOMOLECULES

The Jepsen group[43] also studied the dielectric function of the biomolecules, benzoic acid, and the monosubstituted analogues, salicylic acid (2-hydroxy-benzoic acid), 3- and 4-hydroxybenzoic acid, and aspirin (acetylsalicylic acid) from 0.5 THz to 4.5 THz at 10 K and 300 K using THz-TDS. This series was studied to elucidate the sensitivity of THz vibrational modes to small changes in the overall structure as well as the hydrogen bonding environment.

Despite the similarity of the molecular structure, the absorption spectra shared very little in common. The authors attributed modes below 72 cm^{-1} to lattice vibrations and higher frequency resonances to dimer modes (torsions, out-of-plane bends and asymmetric stretches of the hydrogen bonds) from previous theoretical work.[44] The broadening and redshift of the features with increasing temperature was attributed to mechanical anharmonicity of the potential energy surfaces.

Based on fits of the 10 K data to a frequency dependent complex dielectric function, the center frequencies, line widths, and oscillator strengths (and integrated band intensities) are reported for the three isomers. For the all-trans form, these same parameters are used to model the room temperature data after red-shifting all frequencies by 4.5 cm^{-1} and increasing line widths by a factor of two. Comparison of the absorption features of the three forms enabled some speculation of the types of nuclear motions involved.

Similar spectroscopic studies of small pharmaceutical species have also been conducted using THz-TDS spectroscopy by the Taday group.[45] Investigations of the spectrum below 3 THz were conducted on the active drugs carbamazepine, piroxicam, and theophylline in the solid state and showed that polymorphism and hydration strongly affects the THz spectrum. This and earlier work by the same group at Teraview in the United Kingdom demonstrates that THz methods can be successfully applied to the study of drug species and how subtle hydration, environmental, and structural effects alter drug potency and their controlled manufacture in the pharmaceutical industry.

The impact of mechanical anharmonicity on THz vibrational features was investigated in our group using THz-CW methods by measuring the temperature dependence of the THz absorption spectrum (0.06–4.3 THz) of crystalline biotin.[46,47] The THz spectra obtained at 4.2 K and 298 K are shown in Figure 7.7. A quantitative model was developed to account for observed changes in line shapes and the redistribution of intensity over this temperature range. Such studies are unique to this region because the fundamental vibrational transition frequencies (1 THz $\approx$ 33 cm^{-1}) are comparable to the thermal energy available (kT $\approx$ 200 cm^{-1}) at room temperature. With sufficient spectral resolution (<0.1 cm^{-1}), the enormous change in the level populations that occur over the temperature range from 4.2 K to 298 K may be exploited to recover upper limits on the mechanical anharmonicity of the vibrational modes. The lower part of Figure 7.7 illustrates the impact of temperature on the line shape when (positive) mechanical anharmonicity is introduced at the level of 0.1 percent for a 1 THz vibration.

The model calculations superimposed in Figure 7.7 were obtained as follows. The 4.2 K spectrum was first least-squares fit to Gaussian line shape model containing thirteen lines. The asymmetrically broadened and redshifted features in the room temperature spectrum were then fit using just one adjustable parameter per line, the mechanical anharmonicity. This simple model was shown to explain the intensity redistribution and the asymmetric distortion of the line shape at room temperature. With the further relaxation of the center frequencies in the room temperature fit, the quality of the fit shown in Figure 7.7 was comparable to that obtained at 4.2 K. Because crystal cage dimensions are known to vary by 1 to 3 percent over even smaller temperature ranges,[48–53] such shifts are expected based on current theoretical results where the crystal parameters were fully relaxed.

The best fit Gaussian line widths obtained at 4.2 K are shown in the top panel of Figure 7.8 to increase linearly with frequency. The line shape function and linear increase are both consistent with an inhomogeneous broadening mechanism that likely results from the distribution of crystal lattice defects. Consequently, Lorentzian widths arising from vibrational dephasing (or damping) must be less than these widths enabling conservative lower limits on the vibrational lifetimes of 5 ps to 1 ps. The anharmonicity factors determined for each line are shown in the lower panel of Figure 7.8 as a function of frequency and are all positive indicating the anharmonicity associated with hydrogen bond stretching is more important than negative contributions arising from torsional motions. Moreover, the observed increase with
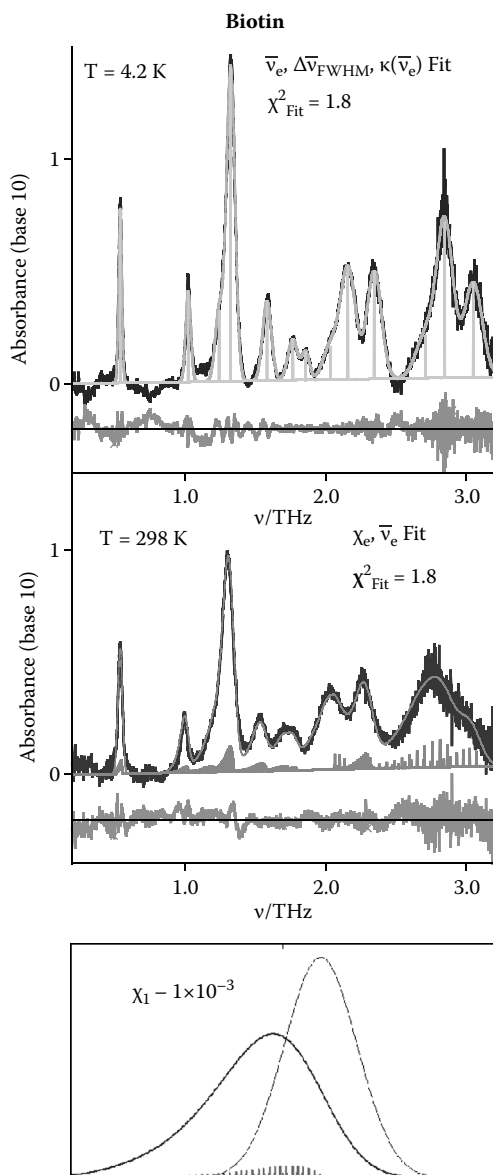
FIGURE 7.7 THz absorption spectra of crystalline biotin obtained at 4.2 K (top panel) and 298 K (middle panel) where the linearly increasing baseline has been removed. Superimposed on the spectra are model predictions with residuals. The absorption features in the 4.2 K spectrum were first identified and fit to a Gaussian lineshape model. The best fit parameters were held fixed in fits to the 298 K spectrum expect for the center frequencies and the anharmonicity factors (χ_e). An example of the impact of anharmonicity ($\chi_e = 0.1$ percent) on the line shape at 1 THz is illustrated at the bottom for the two temperatures.
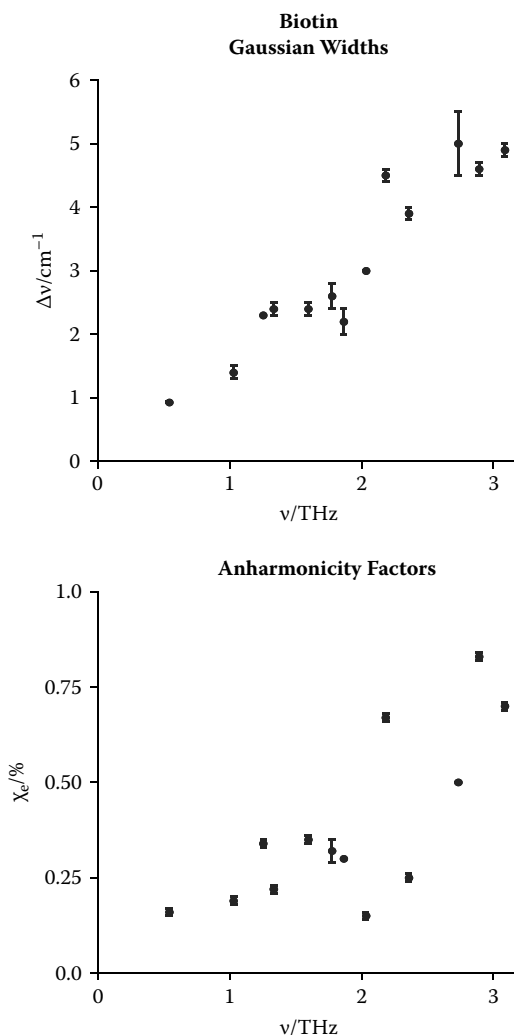
Biotin
Gaussian Widths

Anharmonicity Factors

FIGURE 7.8 Trends observed in the THz absorption spectra of biotin. In the top panel, the increasing Gaussian widths with frequency (obtained from the 4.2 K spectrum) are consistent with an inhomogeneous broadening mechanism. In the lower panel, the increasing anharmonicity factors with frequency (obtained from the 298 K spectrum) suggest the vibrational force fields contain increased contributions from hydrogen bonding stretch potentials.

frequency is as expected because the higher frequency modes are likely to have more hydrogen bonding stretch character. The anharmonic constants represent upper limits since contributions from other line broadening factors would only reduce their magnitude. The significance of this study is that it provides quantitative data to access the impact of anharmonicity on THz spectra and to refine vibrational force field models to account for the temperature dependence of absorption features.

7.6   THZ STUDIES OF LARGE BIOMOLECULES

7.6.1   DNAs AND RNAs

Some of the earliest work on DNA in THz region (0.09–13.5 THz) was performed by Wittlin and coworkers[54] over the temperature range from 5 K to 300 K on highly oriented films of Li-NDA and Na-DNA. Five vibrational modes were identified including the lowest frequency modes at 1.35 THz for Li-DNA and 1.23 THz for Na-DNA. These modes were found to soften with hydration. A simple lattice dynamic model was reasonably successful in explaining the vibrational modes and in the impact of hydration on these modes.

Two artificial RNA single strands composed of polyadenylic acid (*poly-A*) and polycytidylic acid (*poly-C*) were studied by the Jepsen[55] and Tominaga[6] groups using a THz TDS imaging method over the frequency range from 0.1 THz to 4.0 THz ($\Delta\nu_{Inst} \approx 0.5$ cm^{-1}). Their technique was used to record images of spotted arrays on a nonpolar polymer substrate with densities that varied between 2×10^5 g/mole to 7×10^5 g/mol (corresponding to chain lengths of 600–2000 *A* or *C* units). The thin films were formed by dissolving samples in water and by repeated spotting/evaporation of the droplets. A surface profiler was used to character the size of the spots which were 40–50 μm thick with diameters from 1 to 1.5 mm.

Bulk samples in pressed pellet form were first studied and shown to give unstructured absorption that increased linearly between 0.1 THz and 2.5 THz. *Poly-C* was found to absorb more strongly than *poly-A*. The index of refraction of *poly-C* is 10 percent larger than *poly-A*. Analysis on the thin films was performed by taking the difference in integrated absorption between *poly-A* and *poly-C*. Differences of 10 percent for 200-μg samples permitted the easy discrimination between these two forms of RNA. Reliable image contrasts were obtained for spot sizes of 1 mm in diameter and for sample weights of ≈100 μg or more. However, they found little discrimination for sample weights of 20 μg. Furthermore, no sharp spectra features were observed for either bulk samples, freestanding thin films or for the spots. They noted that etalon artifacts were largest when the sample absorption was lowest because of multiple bounces between the interfaces. This study contradicts Globus's work[56] on RNA polymers discussed below where structured absorption is reported. They argue that continuous unstructured absorption is expected in the amorphous condensed phase because of the increasing density of IR-active modes with frequency and the high sensitivity to local environment where additional damping is expected from random interactions with neighboring molecules. They concluded that the lack of distinct spectral features represents a severe limitation for many biologic applications but noted that the careful determination of absorption coefficients could potentially offer additional information about the biomolecule and its environment.

In a series of studies by Globus and coworkers,[57–61] THz spectra of single- and double-stranded DNA and RNA samples have been measured using an FTIR spectrometer typically performed within the frequency range from 10 to 25 cm^{-1} and at a resolution of 0.2–2 cm^{-1}. In the first study, spectra of DNA samples in liquid gel phase were compared with solid film spectra. Thin samples were sealed in polyethylene or polycarbonate films having thicknesses from 10 to 100 μm. These optical pathlengths are smaller than THz wavelengths thereby reducing radiation losses

and minimizing standing wave interference patterns from cell reflections. Polarization properties of absorption were used to characterize samples and many examples were presented to establish the reproducibility of the runs. For liquid gel samples, improved sensitivity from increased fidelity of features (0.3–0.5 cm^{-1}) is reported when compared with the corresponding solid phase spectra. The samples were further characterized using ultraviolet absorption spectra. In these THz studies, discrimination between the native (double-stranded) and denatured (single stranded) forms of DNA was claimed based on changes in four features.

In a second study of DNA, the Globus group[59] investigated the THz spectra of four different types of samples: randomly oriented DNA, partially oriented DNA, artificial poly ribonucleic acid, and *Bacillus subtilis* spores. The four sample types were prepared in various manners to achieve a thin film that was either free-standing or applied to a near-transparent substrate. In general, transmission of samples exceeded 90 percent with structure at the level of 1–2 percent. This apparent structure was found to vary depending upon sample thickness, sample orientation, and water content and interpreted to result from a high density of modes.

Additional DNA studies by Globus and coworkers[61,62] included single- and double-stranded salmon and herring DNA sodium salts, chicken egg ovalbumin, and bacillus subtilis spores. Film samples on nearly transparent substrates were prepared from a water gel in varying concentration and air-dried while liquid samples were sealed between thin polyethylene films. Sample thickness varied between 1 and 250 µm and attempts were made to align the molecules mechanically. It was again noted the extreme sensitivity of the spectra to sample preparation including drying conditions that particularly affect phase transitions of the DNA liquid crystal gels as well as sample thickness, initial concentration in water, and orientation. From reflection and transmission spectra, the frequency-dependent index of refraction and absorption coefficient were determined using an interference spectrometry technique. Transmission spectra of the bacillus subtilis samples appear roughly linear at a level of 88–96 percent across the 15 cm^{-1} range with variations of the order of 1 percent. Results for liquid samples of double-stranded DNA also show roughly linearly decreasing absorption with frequency from 97 to 92 percent across the same range with similar 1 percent variations. Absorption by single-stranded DNA samples[61] was generally 20 percent greater than for double-stranded DNA.

Globus and coworkers[59] also examined the THz spectra of known sequences of single-stranded RNA potassium salts and double-stranded RNA sodium salts. Single strand samples consisted of poly-adenine (A), poly-guanine (G), poly-cytosine (C), and polyuracil (U) and double-strand samples consisted of paired homopolymers C-G and A-U. Films of thickness less than 250 µm were prepared with sample concentration of 2–20 percent in water. Films were either free standing or applied to near-transparent 8 µm thick polycarbonate substrates. Attempts to align the strands were largely unsuccessful and spectral sensitivity to sample preparation, alignment, and interference effects were examined. The frequency-dependent absorption coefficient was calculated for the four single-stranded homopolymers and for the two double-stranded homopolymers, C-G and A-U. Results of theoretical calculations using JUMNA and LIGAND are given using the FLEX force field for hydrogen bonding for RNA fragments.

Terahertz spectroscopy is also sensitive to DNA strand hybridization and has been shown to serve as an analytical tool for DNA sequencing. In several elegant articles by Bolivar and Kurz,[62,63] strong THz difference absorption upon hybridization was demonstrated using THz-TDS spectroscopy, indicating that direct measurement of free strands and hybridized pairs can be readily identified.[62] A novel waveguide approach was further developed, enabling direct hybridization with femtomole sensitivity (comparable to fluorescence tagging methods).[63]

Early work in our group with Roitberg and Markelz[64] reported the THz absorption spectra of calf thymus DNA, bovine serum albumin (BSA; similar to human serum albumen), and type I collagen that have molecular weights of 12 MDa, 66 kDa, and 360 kDa, respectively. Dry samples in polyethylene powder of varying thickness were investigated to access artificial contributions from sample etalon effects. After scaling for molecular path length, samples of different thickness all showed Beer's law behavior. Comparison with undiluted samples showed no measurable difference in the spectra of diluted and undiluted samples indicating that spectral contributions from the instrument could not be resolved by the 0.7 cm^{-1} resolution. Spectral differences in the two systems studied ensured that the observed data were indeed probing biomolecular signatures. For all three biomolecules, broadband absorption spectra devoid of any features were observed at room temperature and for relative humidities (r.h.) of <5 percent. For DNA and BSA, absorbance increased roughly linearly with frequency over the observed range of 0 cm^{-1} to 60 cm^{-1}, whereas absorbance for collagen increased at a faster rate.

Both DNA and BSA were investigated under different conditions of r.h. to study the hydration dependence of the THz response. Pure samples were used to ensure that polyethylene did not hinder sample hydration. Both BSA and DNA exhibited changes in the THz absorbance over the 0 cm^{-1} to 60 cm^{-1} range observed. It was noted that both BSA and DNA are expected to have conformations that depend on the degree of hydration.

DNA has a predominantly helical A conformation at r.h. between 45 percent and 92 percent but a disordered helical A conformation at lower hydration and predominantly helical B conformation at higher hydration. Of the three spectra, the first was taken at <5 percent r.h., the second at 43 percent r.h., which is near the transition point from disordered to ordered helical A state, and the third at 70 percent r.h., where a predominantly ordered helical A conformation is to be expected. Interestingly, the first and last spectra, corresponding to the disordered and ordered helical A conformational states are distinctly different, whereas the second spectrum near the transition point share features from both.

7.6.2 PROTEINS

The conformation of BSA as a function of hydration is not definitively known but is expected to be similar to human serum albumin. The latter has a greater content of β-sheet conformation for r.h. <16 percent but increasing α-helical content for 30 percent to 50 percent r.h. The spectra observed for BSA at <5 percent r.h. is similar to that of denatured BSA but redshifted while the spectrum observed for BSA at 77 percent r.h. shows some differences including an increased redshift.

These differences may reflect the expected changes from β-sheet to α-helix in the conformational composition of the sample.

An important aspect of sample preparation is the time for structural equilibration. Several days may be required for structural equilibrium to be reached after conditions of humidity are changed. The samples studied here, however, were exposed for only 90 minutes before the acquisition of spectra. It is therefore not possible to attribute the changes observed in the THz spectra to the conformation changes expected for equilibrium hydration levels with any certainty. In light of our more recent work on dipeptides[66] and tripeptides,[36] structurally integrated water, non–hydrogen-bonded structural water or superficially adsorbed water can have varying impacts on the THz response.

The three different retinal isomers that occur in the photoactive proteins, rhodopsin, bacteriorhodopsin, and isorhodopsin were investigated by the Jepsen group[65] between 10 cm^{-1} and 100 cm^{-1} at 10 K and room temperature. The three different crystalline forms prepared in polyethylene pellets included the all-trans, 9-cis, and 13-cis isomers that differ only in their conformational structure. The absorption data illustrate the utility of the region for the clear distinction of isomeric samples. Furthermore, the broad absorption features present at room temperature are shown to sharpen and blueshift with decreasing temperature.

The Markelz group has gone on to investigate several other large biomolecules, including hen egg white lysozyme, horse heart myoglobin, and bacteriorhodopsin in collaboration with Whitmire, Hillebrecht, and Birge.[67] Lysozyme and myoglobin powder were diluted with PE powder and pressed into pellet form. Samples were temperature and humidity controlled and spectra were acquired at 77 K and room temperature. For both substances, broadband absorption without identifiable peaks was observed to increase from 5 cm^{-1} to 40 cm^{-1} and then leveled off up to the 80 cm^{-1}. At 77 K, absorption was less than at room temperature across the observed spectrum but remained featureless. The THz spectrum of the organic semiconductor α-hexathiophane in crystalline form was acquired using an FTIR instrument confirm the featureless properties of these samples.

The experimentally measured spectra were compared to the density of states calculated by CHARMm. These calculations were performed on myoglobin and lysozyme structures available from the Protein database without any additional water. The state density for both was roughly 500 modes in the first 80 cm^{-1} with an average separation of 0.2 and 0.16 cm^{-1} for lysozyme and myoglobin, respectively. In both cases, the density of states increased roughly linearly with frequency and leveled off near 60 cm^{-1} in contrast to the experimentally observed maximum absorbance near 40 cm^{-1}. This similarity between the calculated state density and observed THz absorption spectra suggested that all modes are IR active and oscillator strength is frequency independent. However, from CHARMm and quantum chemical calculations on small crystalline systems, unphysical results were found for even these simple systems without a high degree of verisimilitude in the model to mimic the surrounding molecular environment. In particular, the optimized geometry was not consistent with the known geometry without the surrounding molecular environment which, in turn, led to unrealistic outcomes regarding dipole-dipole interaction energies. Furthermore, results employing periodic boundary conditions were

still unsatisfactory. Although the state density found in CHARMm calculations was roughly correct, the positions and intensities of lines could not be reconciled with experimental data.

Markelz in collaboration with Whitmire, Hillebrecht, Birge, and others[68] reported the first demonstration of THz response to conformational state in some interesting and well-conceived studies of bacteriorhodopsin (BR). BR undergoes a light activated photocycle consisting of a series of intermediate conformational states labeled K through O which are all identifiable by their ultraviolet/Vis absorption spectra. The end result of the photoisomerization processes and the interactions between the chromophore and its binding pocket enabled by conformational changes is the transport of a proton from the inner to the outer side of the cell membrane, which generates a potential gradient to drive formation of ATP. In this photocycle, the particular conformation state labeled M has the useful property that its 10 msec lifetime can be prolonged indefinitely at 233 K. Hence, under appropriate conditions of humidity, the M state can be photoinduced at <570 nm and then captured at 233 K to permit the THz spectra to be measured and compared with the spectrum of the ground state conformation.

This work investigated two forms of BR, the wild type (WT) and the D96N mutant, which involves a single residue change. In the D96N mutant, the 96th aspartic acid residue of the 248-residue chain is replaced by asparagine. This seemingly small change of structure results in a 1,000-fold increase in the photocycle time of the molecule.[69] It has been suggested that this increase is due to a loss of conformational flexibility with mutation. Differences in THz response between the WT and mutant BR may therefore suggest a possible correlation with conformational flexibility because motions permitting conformational state transitions may be decomposed into a series of THz vibrational motions. Therefore conformational transitions requiring access to the lowest frequency collective modes may impact the THz response.

The THz absorption spectra of the two types of BR were first obtained at 80 percent r.h. and room temperature for 200-µm thick films on infrasil quartz substrates. In both cases, absorption over the range 0 cm^{-1} to 60 cm^{-1} is broad and increases linearly with frequency but the absorption of the D96N is lower by a factor of 1.8. No difference was observed in spectra taken at 30 percent r.h., but at <5 percent r.h., absorption in both species decreased markedly. The change was too large to be explained by mass-loading effects of water content and may result from intramolecular coupling effects and/or collective modes of interior water clusters.

The normal mode frequencies and eigenvectors derived from CHARMm were used to verify that indeed the very lowest frequency modes are responsible for the largest structural fluctuations while higher frequency modes involve only smaller rms fluctuations. Additional calculations performed on both WT and D96N used an approximation in which the dipole moment squared is proportional to its derivative. Absorption was then estimated using an ad hoc method to account for anharmonicity with no explicit reference to temperature where the sum over states was taken using harmonically approximated states that included only n = 0 to n = 1 transitions. Some agreement was found in comparisons with the initial BR spectra taken at room temperature and 80 percent r.h. However, little difference was found

between the calculated spectra of the WT and mutant BR in contrast to the large differences observed.

In studies comparing illuminated and nonilluminated BR, the THz spectra were first obtained at room temperature for the ground state and then again after cooling to 233 K. For the WT-BR, the broadband featureless absorption decreased markedly with decreasing temperature at the upper end of the 0 cm^{-1} to 60 cm^{-1} range. Conversely, absorption decreased in D96N-BR with decreasing temperature at the lower frequencies in the range. The most striking effect occurred, however, on the photoconversion to the M state at 233 K in which the THz absorption increased relative to that at room temperature. In contrast, the absorption strength remained essentially unchanged in D96N-BR after conversion from the ground state to the M state. On illumination with <480 nm light, the THz response of WT-BR reverted to that previously observed for the ground state at 233 K and at room temperature. A change in M state absorbance was suggested to result from either a change in dipole moment or a change in the distribution of normal modes. Because the former differences are nearly identical for both WT and D96N, changes in state density were suggested to explain the THz response in WT. However, it should be noted that the ultraviolet/Vis spectra used to confirm conformation state indicated a lower M state content for illuminated WT-BR at 233 K than for illuminated D96N-BR at 233 K. Therefore the presence of N or P states in illuminated WT-BR might conceivably be responsible for the observed absorption increase.

The significance of these results is the demonstration of the sensitivity of the THz response to changes in conformation state and to changes of a single residue in the peptide sequence. These studies also raise unanswered questions relating to the quality of theoretical models. The CHARMm results showed no obvious basis for the lower THz absorption of D96N-BR in the ground state compared with WT-BR, nor explained the striking difference in the THz response of WT-BR compared with mutant D96N BR under illumination and conformation change. It is important to note that the CHARMm calculations contained no bound water, whereas more recent studies of crystalline di-[66] and tripeptide[36] systems have shown that the presence of water can strongly affect the THz spectra. These studies have also suggested that the harmonic approximation may not be adequate to reproduce finer details of the observed THz spectra.

The Tominaga group reported the earliest known low frequency THz study of cytochrome C solid in 2002.[6] In more recent work with Chen, Knab, and Cerne, Markelz has studied cytochrome C (CytC), a string of 104 amino acids having a covalently bounded heme group.[69] This electron transfer protein has two forms that are structurally quite similar yet have quite different thermal stability, hydrogen exchange rate, and proteolytic digestion rate. The bound heme group exists in two oxidation states and X-ray measurements, NMR structural measurements and compressibility measurements have indicated that the oxidated state, which is the less stable and more active state, has greater conformational flexibility than the reduced state in spite of their great structural similarity. The THz absorption coefficient of the more flexible, oxidized states is much greater than that of the more stable, less active reduced state. As in the studies of large molecules discussed above, the response is

broad, essentially featureless, and aptly characterized as "glasslike" over the 10 cm^{-1} to 80 cm^{-1} range observed.

In fitting the dielectric response, it was assumed that all modes are active and oscillator strength is more or less uniform across the frequency range of interest. Differences in the fits can be due to differences in dipole coupling strength or state density. The predicted dipole moment difference between the oxidized state and the less flexible reduced state is only about 6 percent. Furthermore, the difference in response is still too large to be explained by increases in state density resulting from the larger water content (and hence system size) of the oxidized state. This outcome is consistent with an increased number of low-frequency modes for the more flexible oxidized state and illustrates the use of THz spectra as a measure of the accessibility of the low frequency modes crucial to conformational flexibility.

In the most recent work from the Markelz group,[70] the THz response of hen egg white lysozyme is investigated as a function of hydration for r.h. from 3 percent to 39 percent. Other workers using NMR, specific heat and microwave spectroscopy measurements have previously identified a transition point near 27 percent r.h. at which the first hydration shell is filled and bulk water begins to accumulate. Evidence of this transition is also seen in the THz spectra in this work for which very uniform thin films (≈100 μm thick) of lysozyme were prepared on quartz substrates. The lysozyme samples were allowed to reach equilibrium in a humidity-controlled cell at room temperature prior to obtaining spectra.

For lysozyme, the absorption coefficient α increases nearly linearly with frequency, and the index of refraction n is nearly flat for all levels of hydration shown. Of particular interest is the sudden increase in the entire absorption curve at the 27 percent r.h. and the subsequent shifts which are even larger as hydration is increased. A similar effect is seen for index n at the same transition point. The apparent discontinuity is indicative of the transition from occupation of bound water sites to bulk water accumulation.

In previous work, attempts were made to model large biomolecules using CHARMm to obtain the normal mode distribution and then with simplifying assumptions about the dipole interactions to estimate the THz absorption spectrum. Such models were not especially successful and, in this work, a new approach was tried. Because the system at room temperature is believed to be composed of multiple conformations, one approach is to consider the conformational lifetime rather than determining the normal mode distribution of each particular conformation. Several such models were investigated but results were not entirely satisfactory. Simple one- or two-relaxation time models gave reasonably good fits, but unfortunately unrealistically short lifetimes as compared with the relaxation times found experimentally by others. On the other hand, good fits can be found by adding a sufficient number of oscillators and relaxation times but no physical model justifies such an approach and so little is gained.

7.6.3 POLYSACCHARIDES

Investigation of the THz spectroscopy of polysaccharides has also been performed in our group (unpublished). Although only a few studies have been conducted, the
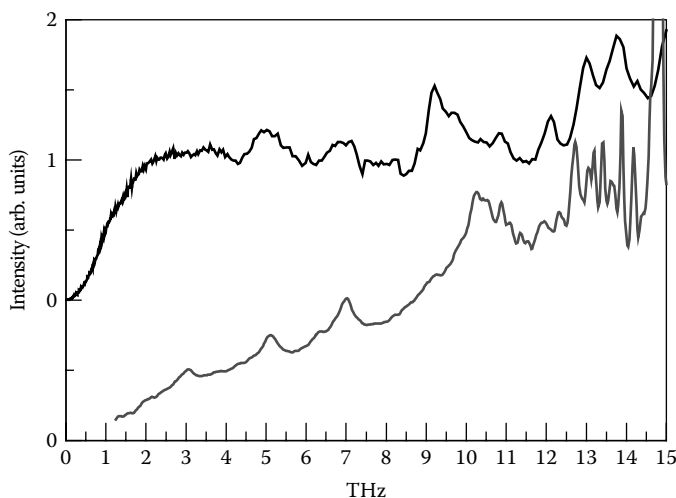
FIGURE 7.9 Inelastic neutron spectrum (top) and THz absorption spectrum (bottom) for dispersed cellulose powder in polyethylene matrix. The triple-helix polymer exhibits distinct positive absorption features at 1, 3, 5, 7, 9, and higher frequencies, presumably arising from internal collective motions of the helix.

infrared absorption and inelastic neutron spectra of simple purified cellulose powder have been obtained (Figure 7.9). Cellulose is a hydrogen-bonded triple helix, where each strand is composed of polyglucose units. There are clear absorption and scattering features that occur at 3, 5, 7, 9, and 11 THz as shown in the figure. Although there are no modeling results for this system, it is surmised that the periodicity of these systems probably produces interstrand twisting or stretching motions that show up as harmonics in the THz spectrum. Further studies of modified cellulose (different crystalline forms or substituted species such as nitrocellulose) are clearly warranted, as are periodic boundary CPMD or related solid-state calculations to help identify the origin of these unique polymeric spectral features.

7.7   THZ STUDIES OF BIOMOLECULES IN LIQUID WATER

Two recent studies published at nearly the same time have addressed the use of THz spectroscopy to probe biomolecules in aqueous environments. For example, the Havenith group[71] found that the THz region provides a sensitive probe of solute-induced changes in water's hydrogen bonding network near the vicinity of a biomolecule. The simple sugar, lactose, has been investigated using p-Ge-laser THz source[72] over the spectral region from 2.3 THz and 2.9 THz. Highly accurate absorption measurements were performed on bulk water and lactose power and compared with total absorption measured for a series of different concentrations of lactose in water. The challenge of these measurements arises because the absorption coefficient of water is 450 cm^{-1} at 3 THz and therefore measurements were performed in a cell 150 μm in length. Absorption for solutions was found to be greater than the fractional sum

(by volume) of the bulk absorptions. This amounted to 65 percent of the total absorption for 0.7 mol/L solution! Further measurements of volume and modeling of hydration shells indicated the solute induced perturbations extended more than 5.13[12] Å from the surface of lactose. This volume corresponds to 123 water molecules/lactose. Simulations using molecular mechanics (and the CHARMm force field) reveal a slowing down of the hydrogen bond rearrangement dynamics in the hydration spheres. This study demonstrates that THz spectroscopy probes the coherent solvent dynamics on the subpicosecond timescale and gives quantitative measure of the extent the solute influences biological water (i.e., water near the surface of a biomolecule).

In a THz study (0.1–1.2 THz) of myoglobin (Mb) using THz-TDS, Durbin and Zhang[73] reported an increase in the absorption per protein molecule in the presence of "biologic" water. Absorption measurements were made on samples between 3.6 wt percent to 90 wt percent water[74] as confirmed by weight loss on heating and accurate optical pathlength determinations critical to this analysis were determined from comparisons made using 100 μm and 200 μm cuvettes. As with the other large biomolecules described previously, nearly continuous broadband absorption that increased linearly with frequency was observed. However, deviations of the normalized absorption from an idealized noninteracting Mb-water model showed enhanced THz absorption rather than a reduction expected on theoretical grounds from a decrease in the 2.5 Debye moment of bulk water to 0.8 Debye because of hindered motions experienced by biologic water. The increased absorption and hence, polarizability of Mb across the THz region (especially above 90 wt percent water), is suggested to result from the enhanced mobility of the side chains along the protein[75] and for more than 90 wt percent water, increased rotational freedom of Mb that occurs when the average separation (>6 nm) becomes significantly larger than radius of gyration (1.5 nm). The significance of these latter two studies is that through quantitative THz absorption measurements, the solvent-biomolecule interfacial structure and dynamics may be examined directly.

7.8 CONCLUSIONS AND FUTURE INVESTIGATIONS

Application of terahertz spectroscopic methods to the study of important biological species, ranging from simple sugars to complex DNAs has been reviewed. We have tried to give the reader a glimpse at the different types of systems investigated to date, and hope that there is ample evidence that the investigations performed have revealed new and exciting properties of these systems. IR methods have improved a great deal in the last decade, so novel studies of complex systems can now be performed. We apologize for not including references to all studies and investigators involved in this rapidly growing field. Although there are many hurdles to overcome, investigations of complex biomolecular species in the solid-state are under way and many new groups are encouraged to become involved in this growing research area.

Interpretation of resultant THz spectra absolutely requires theoretical modeling to deduce the origin of spectral absorptions and perform intra- and intermolecular vibrational mode assignments. At present, some approaches are adequate and yield quite good agreement with experimental spectra. These methods are also improving with the availability of more advanced computational methods and sophisticated

hardware. Feedback between experiment and modified theoretical approaches will continue to advance our ability to interpret and extract structural details of larger molecular systems from THz spectra. This interplay between experimentalists and theorists will hopefully produce better force fields and computational methods to help predict and confirm spectral assignments for high molecular weight biomolecules.

Finally, we might suggest that time-resolved investigations of biomolecular dynamics and function in aqueous environments are the ultimate goal for experimentalists. The challenge of probing biomolecule spectra in a highly absorbing media is great, and the influence of hydrogen-bonding on solutes by water of hydration is unknown, but recent studies of species in water give us hope that this goal will one day be achieved.

ACKNOWLEDGMENTS

REFERENCES

1. R. H. Callender, R. B. Dyer, R. Gilmanshin, W. H. Woodruff, Fast events in protein folding: the time evolution of primary processes. *Annu. Rev. Phys. Chem.*, 49, (1998) 173.
2. W. A. Eaton, V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, J. Hofrichter, Fast kinetics and mechanisms in protein folding. *Ann. Rev. Biophys. Biomolecular Structure*, 29 (2000) 327.
3. S. Takahashi, S. R. Yeh, T. K. Das, C. K. Chan, D. S. Gottfried, D. L. Rousseau, *Nature Struc. Biol.*, 4 (1997) 44.
4. A. G. Markelz, A. Roitberg, E. J. Heilweil, *Chem. Phys. Lett.*, 320, (2000) 42.
5. R. Nossal and H. Lecar, *Molecular and cell biophysics*, Addison-Wesley, Redwood City, Ca, 1991.
6. K. Yamamoto, K. Tominaga, H. Sasakawa, A. Tamura, H. Murakami, H. Ohtake, N. Sarukura, *Bulletin of the Chemical Society of Japan*, 75 (2002) 1083.
7. B. S. Hudson, *J. Phys. Chem. A*, 105 (2001) 3949.
8. J. T. Kindt, C. A. Schmuttenmaer, *J. Phys. Chem.*, 100 (1996) 10373.
9. Certain commercial equipment, instruments, or materials are identified in this article to adequately specify the experimental procedure. In no case does identification imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.
10. M. C. Beard, G. M. Turner, C. A. Schmuttenmaer, Progress towards two-dimensional biomedical imaging with THz spectroscopy. *J. Phys. Chem. B.* 106 (2002) 7146.
11. A. S. Pine, R. D. Suenram, E. R. Brown, K. A. McIntosh, *J. Mol. Spectro.* 175, (1996) 37.
12. K. A. McIntosh, E. R. Brown, K. B. Nichols, O. B. McMahon, W. F. DiNatale, T. M. Lyszczarz, *Appl. Phys. Lett.*, 67 (1995), 3844.
13. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, et al., Gaussian 98 (Revision A.11.3), Gaussian, Inc., Pittsburgh PA, 2001.

AU: Please provide article title

AU: Please provide article title

AU: Please provide article title

AU: Please provide article title

AU: Please provide article title

AU: Please provide article title

AU: Please provide article title

AU: Please provide title

14. See: http://webbook.nist.gov/chemistry
15. A. D. MacKerell, N. Banavali, Development and current status of the CHARMM force field for nucleic acids. N. Foloppe *Biopolymers* 56 (2000) 257.
16. R. B. Gerber, C. M. Chaban, S. K. Gregurick, B. Brauer, Vibrational spectroscopy and the development of new force fields for biological molecules. *Biopolymers*, 68(3) (2003), 370.
17. S. K. Gregurick, S. A. Kafafi, *J. Carbohydrate Chemistry*, 18 (1999) 867.
18. M. M. Harding, H. A. Long, *Acta Cryst.* B24 (1968) 1096.
19. B. Khawas, *Acta Cryst.* B27 (1971) 1517.
20. T. J. Kistenmacher, G. A. Rand, R. E. Marsh, *Acta Cryst.* B30 (1974) 2573.
21. C. H. Görbitz, B. Dalhus, *Acta Cryst.* C52 (1996) 1756.
22. M. N. Frey, M. S. Lehman, T. F. Koetzle, W. C. Hamilton, *Acta Cryst.* B29 (1973) 876.
23. K. A. Kerr, J. P. Ashmore, T. F. Koetzle, *Acta Cryst.* B31 (1975) 2022.
24. R. Car, M. Parrinello, *Phys. Rev. Lett.* 55 (1985) 2471.
25. N. Trouiller, J. L. Martins, *Phys. Rev.* B 43 (1991) 1943.
26. L. Kleinman, D. M. Bylander, *Phys. Rev. Lett.* 4, (1982) 1425.
27. R. Resta, *Phys. Rev. Lett.*, 80 (1998) 800.
28. D. A. McQuarrie, Statistical mechanics. Harper-Collins Publishers: New York, 1976.
29. B. Delley, *J. Chem. Phys.* 113 (2000) 7756.
30. J. P. Perdew, Y. Wang, *Phys. Rev. B*, **45** (1992) 13244.
31. J. Higgins, X. Zhou and R. Lin, *Spectrochim. Acta A.*, 53 (1997) 721.
32. J. Melinger, N. Laman, S. Sree Harsha, and D. Grischkowsky, *Appl. Phys. Lett.*, 89 (2006) 251110.
33. M. Kutteruf, C. Brown, L. Iwaki, M. Campbell, T. A. Korter, E. J. Heilweil, *Chem. Phys. Lett.*, 375, (2003) 337.
34. T. M. Korter, R. Balu, M. B. Campbell, M. C. Beard, S. K. Gregurick, E. J. Heilweil, *Chem. Phys. Lett.*, 418, (2005) 65–70.
35. R. Balu, S. K. Gregurick, E. J. Heilweil, Determination of enantiomeric composition by terahertz spectroscopy: mixtures of D- and L-tryptophan. In preparation.
36. K. Siegrist, C. R. Bucher, I. Mandelbaum, A. R. Hight Walker, R. Balu, S. K. Gregurick, D. F. Plusquellic, *J. Am. Chem. Soc.*, 128, (2006) 5764.
37. M. Yamaguchi, K. Yamamoto, M. Tani and M. Hangyo, IEEE conference proceedings on biological and medical applications, ultrafast chemistry and physics, 13th Int. Conf. on Infrared and Millimeter Waves, paper WB3-3, p. 477.
38. A. Hempel, N. Camerman, A. Camerman, Biopolymers, 31, (1991) 187; J. K. Fawcett, N. Camerman, A. Camerman, *Acta. Cryst. B.* 31 (1975) 658.
39. W. Qian, J. Bandekar, S. Krimm, *Biopolymers*, 31, (1991) 193.
40. Z. Bihary, R. B. Gerber, *J. Chem. Phys.* 115, (2001) 2695.
41. B. M. Fischer, M. Walther and P. U. Jepsen, *Phys. Med. Biol*, 47, (2002) 3807.
42. M. Walther, B. M. Fischer, and P. Uhd Jepsen, *Chem. Phys.*, 288, (2003) 261.
43. M. Walther, P. Plochocka, B. M. Fischer, H. Helm and P. U. Jepsen, *Biopolymers* (Biospectrscopy), (2002) 310.
44. H. R. Zelsmann, Z. Mielke, *Chem. Phys. Lett.* 186, (1991) 501.
45. J. A. Zeitler, K. Kogermann, J.antanen, T. Rades, P. F. Taday, M. Pepper, J. Aaltonen, C. J. Strachan, Int. J. Pharmaceuticals, 334, 78 (2007).
46. D. F. Plusquellic, T. M. Korter, G. T. Fraser, R. J. Lavrich, E. C. Benck, C. R. Bucher, J. Domench, A. R. Hight Walker, in *Terahertz sensing technology. Volume 2:*

AU: Please provide article title for all references

AU: Please provide updated publication information

Au: are these two different articles cited in one reference? Please separate

emerging scientific applications and novel device concepts, (World Scientific, 2003, ed. D. L. Wollard, W. R. Loerop, and M.S. Shur), 13(4), (2003) 385.

47. T. M. Korter and D. F. Plusquellic, *Chem. Phys. Lett.* 385, (2004) 45.
48. B. Dalhus, C. H. Gorbitz, *Acta. Chrys. C*, Vol. 52, (1996) 1759.
49. Dawson and Mathieson, *Acta Chrys.* 4, (1951) 475.
50. W.-Q. Wan, Y. Gong, Z.-M. Wang, C.-H. Yan, *Chinese J. Struct. Chem.* 22, (2003) 539.
51. B. Dalhus, C. H. Gorbitz, *Acta. Chem. Scand.* 50, (1996) 544.
52. W. Q. Wang,Y. Gong, Z. Liang, F. L. Sun, D. X. Shi, H. J. Gao, X. Lin, P. Jiang, Z. M. Wang, *Surf. Sci.* 512, (2002) 379.
53. K. Tori, Y. Iitaka, *Acta Cryst*a. B 26, (1970) 1317.
54. A. Wittlin, L. Genzel, F. Kremer, S. Haseler, A. Poglitsch, *Phys. Rev.* A, 34 (1986) 493.
55. B. M. Fischer, M. Hoffmann, H. Helm, R. Wilk, F. Rutz, T. Kleine-Ostmann, M. Koch, P. U. Jepsen, *Optics Express* 13, (2005) 5205.
56. T. W. Crowe, T. Khromova, B. Glemont, J. Hesler, *J. Phys. D: Appl. Phys.* 39, (2006) 3405.
57. T. Globus, D. Wollard, T. W. Crowe, T. Khromova, B. Gelmont and J. Hesler, *J. Phys. D* 39 (2006) 3405.
58. T. R. Globus, D. L. Woolard, A. C. Samuels, B. L. Gelmont, J. Hesler, T. W. Crowe, M. Bykhovskaia, *J. App. Phys*., 91 (2002) 6105.
59. T. Globus, M. Bykhovskaia, D. L. Woolard, B. Gelmont, *J. Phys. D* 91 (2003) 1314.
60. T. Globus, R. Parthasarathy, T. Khromova, D. Woolard, N. Swami, A. J. Gatesman, J. Waldman, *Proc. SPIE*, 5584 (2004).
61. R. Parthasarathy, T. Globus, T. Khromova, N. Swami, D. Woolard, *App. Phys. Lett.* 87 (2005) 113901.
62. M. Brucherseifer, M. Nagel, P. H. Bolivar, H. Kurz, *Appl. Phys. Lett*., 77, (2000) 4049.
63. M. Nagel, F. Richter, P. H. Bolivar, H. Kurz, *Phys. Med. Biol*., 48 (2003) 3625.
64. A. G. Markelz, A. Roitberg, E. J. Heilweil, *Chem. Phys. Lett.* 320, 42–48 (2000).
65. M. Walther, B. Fischer, M. Schall, H. Helm, P. Uhd Jepsen, *Chem. Phys Lett.* 332, (2000) 389.
66. K. Siegrist, D. F. Plusquellic, CW THz studies of the crystalline dipeptide nanotubes. In preparation.
67. A. Markelz, S. Whitmire, J. Hillebrecht, R. Birge, *Phys. Med. Biol.* 47, (2002) 3797.
68. S. E. Whitmire, D. Wolpert, A. G. Markelz, J. R. Hillebrecht, J. Galan, R. R. Birge, *Biophys. J.* 85, (2003) 1269.
69. J.-Y. Chen, J. R. Knab, J. Cerne, A. Markelz, Phys. Rev. E. 72, (2005) 040901R.
70. J. Knab, J.-Y. Chen, A. Markelz, *Biophys. J.* 90, (2006) 2576–2581.
71. U. Heugen, G. Schwaab, E. Bründermann, M. Heyden, X. Yu, D. M. Leitner, and M. Havenith, PNAS, 103, 12301 (2006).
72. E. Bründermann, D. R. Chamberlin, E. E. Haller, *Appl. Phys. Lett.* 73 (1998) 2757.
73. C. Zhang, S. M. Durbin, *J. Phys. Chem. B.* 110 (2006) 23607.
74. C. Zhang, E. Tarhan, A. K. Ramdas, A. M. Weiner, S. M. Durbin, *J. Phys. Chem. B.* 108 (2004) 10077.
75. S. Bone, R. Pethig, J. Mol. Biol. 157 (1982) 157; S. Bone, R. Pethig, *J. Mol. Biol.* 181 (1985) 323.

AU: Please provide authors' initials

AU: Please provide updated publication information